On common invariant cones for families of matrices☆

Leiba Rodman[a,*], Hakan Seyalioglu[b], Ilya M. Spitkovsky[a]

[a] *Department of Mathematics, College of William and Mary, Williamsburg, VA 23187-8795, USA*

[b] *Department of Mathematics, UCLA, Los Angeles, CA 90024, USA*

Abstract

The existence and construction of common invariant cones for families of real matrices is considered. The complete results are obtained for 2×2 matrices (with no additional restrictions) and for families of simultaneously diagonalizable matrices of any size. Families of matrices with a shared dominant eigenvector are considered under some additional conditions.

Key words: Invariant cones, common invariant cones, Vandergraft matrices

2000 MSC: 15A48

1. Introduction

The theory of nonnegative matrices, and more generally of matrices that leave invariant a convex, closed, pointed, solid cone, is classical; we mention here the books [1, 2] among many others; see also [3] for a review of many results, including recent ones, and extensive bibliography. More generally, real matrices that leave invariant a convex, closed, pointed, solid cone, have been studied in [4, 5, 6, 7, 8, 9]. A complete characterization of such matrices in terms of spectral structure was obtained in [5]. An interesting application to the multiple agents randezvous problem is given in [10].

☆Research of all authors was supported in part by NSF grant DMS-0456625. Research of LR was supported in part also by a Summer Research Grant from the College of William and Mary.

*Corresponding author

Email addresses: lxrodm@math.wm.edu (Leiba Rodman), hseyalioglu@ucla.edu (Hakan Seyalioglu), ilya@math.wm.edu, imspitkovsky@gmail.com (Ilya M. Spitkovsky)

Recently, several works appeared studying matrices having common invariant convex, closed, pointed, solid cones. These works have been motivated primarily by applications in Glass networks [11] and joint spectral radius [12, Theorem 1]. Glass networks are continuous-time switching networks used to model gene regulatory networks and neural networks; see [11] and references there for an in depth discussion on Glass networks.

The paper [11] actually served as a motivation for the current paper. We develop here results on matrices having common invariant cones. The auxiliary Section 2 contains necessary notions and definitions, in particular that of a proper cone and a dominant eigenvector. In Section 4, a full description is given of families of 2×2 real matrices having common invariant proper cones. As it turns out even in this case the characterizations are rather involved, and the proofs not immediate. Some partial results (for pairs of diagonalizable but not simultaneously reducible matrices) in this venue were obtained in [11]. Our approach is based on the description of all invariant cones for a single 2×2 matrix given in Section 3. In spite of its elementary nature, we did not find this description in the literature, and include it for the sake of self containment. Section 5 contains the existence criterion for (and actually a construction of) a common invariant cone of a family of simultaneously diagonalizable matrices, while Section 6 provides some sufficient conditions for such a cone to exist when the matrices share the dominant eigenvector. Finally, Section 7 consists of several examples illustrating both the results obtained and their limitations.

2. Preliminaries and definitions

Let $\mathbb{R}$ be the field of real numbers, $\mathbb{R}^n$ the set of real n-component column vectors, and $\mathbb{R}^{m\times n}$ the set of real $m \times n$ matrices. *All matrices in the present paper are assumed to be real, unless explicitly stated otherwise.* A set $\mathcal{K} \subseteq \mathbb{R}^n$ is a *cone* if $a\mathcal{K} \subseteq \mathcal{K}$ for all scalar multiples $a \geq 0$. A cone $\mathcal{K}$ is said to be *proper* if $\mathcal{K} + \mathcal{K} \subseteq \mathcal{K}$ (so that $\mathcal{K}$ is *convex*), *closed*, *pointed* ($\mathcal{K} \cap -\mathcal{K} = \{0\}$) and *solid* (the interior of $\mathcal{K}$ is nonempty).

For X being a subset of $\mathbb{R}^n$ or $\mathbb{R}^{m\times n}$, we denote by $\operatorname{Cone} X$ the smallest convex cone containing X and say that X *generates* $\operatorname{Cone} X$. Of course, $\operatorname{Cone} X$ is nothing but the set of all (finite) linear combinations of elements of X with non-negative coefficients. A cone having a finite generating set is called *polyhedral*. Polyhedral cones are always closed.

For a square matrix A, by the *degree* of its eigenvalue λ in this paper we understand its multiplicity as a root of the minimal polynomial of A (that is, the size of the largest block, in the Jordan canonical form of the matrix, corresponding to the eigenvalue λ). We will denote the eigenvalues of an $n \times n$ matrix A by $\lambda_1(A), \ldots, \lambda_n(A)$ (or simply by $\lambda_1, \ldots, \lambda_n$ if the choice of the matrix is clear from the context), always taking $\rho(A) = \lambda_1$ provided that the *spectral radius* $\rho(A)$ of A is an eigenvalue. We will call the respective eigenvector (eigenspace) the *dominant eigenvector* (resp., *dominant eigenspace*) of A. In case when an eigenspace is one dimensional, we will (naturally) call it an *eigenline*. We will also use the term *eigenray* for each of the two rays into which an eigenline is partitioned by the origin. Finally $\sigma(A)$ will be used to denote the set of all eigenvalues of A.

A cone $\mathcal{K} \subseteq \mathbb{R}^n$ is said to be *invariant* under $A \in \mathbb{R}^{n\times n}$ if $Ax \in \mathcal{K}$ for every $x \in \mathcal{K}$. The following remark is trivial, but will be useful in our analysis.

Remark 1. *A cone $\mathcal{K} = \mathrm{Cone}\,\{v_1, \ldots, v_m\}$ is A-invariant if and only if $Av_j \in \mathcal{K}$ for $j = 1, 2, \ldots, m$.*

The following result was proved by Vandergraft [5].

Theorem 1. *$A \in \mathbb{R}^{n\times n}$ has an invariant proper cone if and only if*

(i) *The spectral radius $\rho(A) \in \sigma(A)$, and*

(ii) $\deg \lambda_1(A) \geq \deg \lambda_i(A)$ *for every eigenvalue $\lambda_i(A)$ with $|\lambda_i(A)| = \lambda_1(A)$.*

If conditions (i)-(ii) *hold, then also*

(iii) *Any A-invariant proper cone contains a dominant eigenvector of A.*

For spectral criteria for existence of polyhedral proper invariant cones see [6, 9].

We will be using the term *Vandergraft matrices* for real matrices satisfying conditions (i) and (ii) of Theorem 1, denoting the set of all such $n\times n$ matrices by $\mathbb{R}_V^{n\times n}$.

3. Invariant proper cones for 2 × 2 matrices

It is very easy to characterize matrices in $\mathbb{R}_V^{2\times 2}$. Namely, condition (i) of Theorem 1 is equivalent to

$$(\mathrm{trace}\, A)^2 \geq 4 \det A, \quad \mathrm{trace}\, A \geq 0, \tag{1}$$

the first inequality in (1) meaning simply that the eigenvalues of A are real while the second inequality guarantees that the one with the bigger absolute value is non-negative. Since condition (ii) then holds automatically, a 2×2 matrix A is Vandergraft if and only if it satisfies (1).

Conditions (1) hold, in particular, when both eigenvalues λ_1, λ_2 of A are non-negative. Description of all A-invariant proper cones in this case is given by the following two theorems, dealing with diagonalizable and non-diagonalizable matrices A separately. Of course, in the former situation only the case $\lambda_1 \neq \lambda_2$ is of interest, because otherwise A is a scalar matrix which leaves every cone invariant.

Theorem 2. *Let a 2×2 matrix A be diagonalizable, with $\lambda_1 > \lambda_2 \geq 0$. Then a proper cone $\mathcal{K} \subset \mathbb{R}^2$ is A-invariant if and only if it contains an eigenvector of A corresponding to λ_1 and its interior does not intersect the eigenline of A corresponding to λ_2.*

Proof. "Only if" part. An A-invariant proper cone $\mathcal{K}$ must contain an eigenvector of A corresponding to λ_1, as follows from Theorem 1, part (iii). Denote this vector by u_1 and suppose for a moment that there is an eigenvector u_2 of A corresponding to the eigenvalue λ_2 and lying in the interior of $\mathcal{K}$. Then for sufficiently large $M > 0$ also $-u_1 + Mu_2 \in \mathcal{K}$, and for all $n = 1, 2, \ldots$,

$$(\lambda_1^{-1}A)^n(-u_1 + Mu_2) = -u_1 + M(\lambda_2/\lambda_1)^n u_2 \in \mathcal{K}.$$

Letting $n \to \infty$, from the closedness of $\mathcal{K}$ we conclude that $-u_1 \in \mathcal{K}$. This, however, contradicts pointedness of $\mathcal{K}$.

"If" part. Any proper cone in $\mathbb{R}^2$ is generated by two linearly independent vectors: $A = \operatorname{Cone}\{v_1, v_2\}$. The conditions imposed on $\mathcal{K}$ mean that, after appropriate scalings, its generating vectors can be written as

$$v_1 = u_1 + u_2, \quad v_2 = u_1 - xu_2,$$

where $x \geq 0$. (Here u_1, u_2 are eigenvectors corresponding to λ_1, λ_2, respectively.) Then

$$Av_1 (= \lambda_1 u_1 + \lambda_2 u_2) = \frac{x\lambda_1 + \lambda_2}{1 + x} v_1 + \frac{\lambda_1 - \lambda_2}{1 + x} v_2 \in \mathcal{K}$$

and

$$Av_2 (= \lambda_1 u_1 - x\lambda_2 u_2) = \frac{x(\lambda_1 - \lambda_2)}{1 + x} v_1 + \frac{\lambda_1 + x\lambda_2}{1 + x} v_2 \in \mathcal{K}.$$

A-invariance of $\mathcal{K}$ therefore follows from Remark 1. $\square$

Let now $A \in \mathbb{R}^{2\times 2}$ be non-diagonalizable. Then, for any $v \in \mathbb{R}^2$,

$$Av = \lambda v + xu, \tag{2}$$

where λ is the eigenvalue of A, u is its (arbitrarily fixed) eigenvector, and $x \in \mathbb{R}$. We will say that v is *positively/negatively associated* with u (*relative to* A, if there is a need to mention the matrix explicitly) if in (2) $\pm x > 0$. Observe that $x = 0$ if and only if v belongs to the eigenline of A, that is, is a scalar multiple of u.

Of course, v is positively associated with u if and only if $-v$ is negatively associated with u if and only if $-v$ is positively associated with $-u$. Geometrically speaking, the plane $\mathbb{R}^2$ is partitioned by the eigenline of A into two open half-planes; one consisting of vectors positively associated with u, and the other of vectors negatively associated with u.

Theorem 3. *Let $A \in \mathbb{R}^{2\times 2}$ be a non-diagonalizable matrix with the eigenvalue $\lambda \geq 0$. Then a proper cone $\mathcal{K}$ is A-invariant if and only if it is given by $\mathcal{K} = \mathrm{Cone}\{u, v\}$, where u is an eigenvector of A and v is positively associated with u relative to A.*

Proof. "If" part. Since $\lambda \geq 0$, from (2) it follows that $Av \in \mathrm{Cone}\{u, v\}$, because $x \geq 0$. Obviously, $Au = \lambda u$ also lies in $\mathrm{Cone}\{u, v\}$. The desired result now follows from Remark 1.

"Only if" part. Let a proper cone $\mathcal{K}$ be A-invariant. Due to Theorem 1(iii), there is an eigenvector of A lying in $\mathcal{K}$. Denoting it by u, observe that vectors negatively associated with u cannot lie in $\mathcal{K}$. Indeed, if $\lambda = 0$ and (2) holds with $x < 0$, then

$$v \in \mathcal{K} \Longrightarrow -u \in \mathcal{K},$$

which contradicts the pointedness of $\mathcal{K}$. For $\lambda > 0$, (2) implies

$$A^n v = \lambda^n v + nx\lambda^{n-1}u, \quad n = 1, 2, \dots.$$

Consequently, if $v \in \mathcal{K}$ and $x < 0$, then

$$-u = \lim_{n\to\infty} \frac{1}{n\,|x|}\lambda^{1-n}A^n v \in \mathcal{K}$$

— once again, a contradiction with the pointedness of $\mathcal{K}$.

Since in every neighborhood of u there are vectors negatively associated with it, u cannot lie in the interior of $\mathcal{K}$. Thus, it must be one of its generating vectors. The other generating vector v, being linearly independent with u, must be positively associated with it. So, $\mathcal{K}$ indeed is of the desired form. $\square$

Corollary 1. *For non-diagonalizable Vandergraft 2×2 matrices, the dominant eigenvector lies on the boundary of their invariant proper cones.*

As follows from Theorem 2, for diagonalizable 2×2 matrices with positive eigenvalues the dominant vector can lie both in the interior and on the boundary of their invariant cones.

We turn now to the remaining case of matrices A with negative determinants. Denote the eigenvalues of A by $\lambda_1(> 0)$ and $\lambda_2(< 0)$, and let u_1, u_2 stand for the respective eigenvectors.

Theorem 4. *Let $A \in \mathbb{R}^{2\times 2}$ and $\det A < 0$. Then a proper cone $\mathcal{K} \subset \mathbb{R}^2$ is A-invariant if and only if it can be represented as $\mathcal{K} = \mathrm{Cone}\{v_1, v_2\}$, where*

$$v_j = u_1 + c_j u_2 \quad (j = 1, 2) \tag{3}$$

and

$$c_1 > 0,\ c_2 < 0, \quad \frac{\lambda_1}{\lambda_2} \le \frac{c_1}{c_2} \le \frac{\lambda_2}{\lambda_1}. \tag{4}$$

Proof. An A-invariant pointed cone cannot contain eigenvectors of A corresponding to a negative eigenvalue. Thus, all vectors $v \in \mathcal{K}$ (in particular, its generators) in their expansion along the eigenbasis $\{u_1, u_2\}$ have the same sign coefficients corresponding to u_1. Switching from u_1 to $-u_1$ if needed, we may without loss of generality suppose that these coefficients are positive. Scaling v_1 and v_2 if necessary, we arrive at (3). Yet another change (from u_2 to $-u_2$, or flipping v_1 with v_2) allows us without loss of generality suppose that $c_1 > c_2$.

On the other hand, for v_j given by (3) we have

$$Av_j = \lambda_1(u_1 + \frac{\lambda_2}{\lambda_1} c_j u_2).$$

Consequently, Av_j lie in the cone $\mathcal{K}$ if and only if the numbers $\lambda_2 \lambda_1^{-1} c_j$ lie in $[c_2, c_1]$. This is equivalent to (4). $\square$

Corollary 2. *Let A be a 2×2 Vandergraft matrix with negative determinant. Then any dominant eigenvector of A lies in the interior of those A-invariant proper cones that contain the eigenvector.*

Note that conditions (4) are consistent if and only if $\det A < 0$ and $\operatorname{trace} A \ge 0$, which of course agrees with (1). If this is indeed the case, for

every non-zero vector v different from the eigenvectors of A there exist A-invariant proper cones $\mathcal{K}$ with v being one of the generators. The second generators of these cones form yet another convex cone, described by (4) with one of c_j being determined by v and the other serving as a parameter. The latter cone degenerates into a single ray if and only if $\operatorname{trace} A = 0$ (equivalently: A^2 is a scalar multiple of the identity), when necessarily $c_1 = -c_2$.

It is very easy to produce directly an A-invariant cone with arbitrarily chosen generator v for any 2×2 matrices A with

$$\det A \leq 0, \quad \operatorname{trace} A \geq 0. \tag{5}$$

Lemma 5. *Let $A \in \mathbb{R}^{2\times 2}$ satisfy* (5). *Then $\mathcal{K}_v := \operatorname{Cone}\{v, Av\}$ is A-invariant for any $v \in \mathbb{R}^2$, $v \neq 0$.*

Of course, $\mathcal{K}_v$ is proper if and only if v is not an eigenvector of A.

Proof. Indeed, $\mathcal{K}_v$ is generated by v and Av. The first of these generators is mapped by A into $\mathcal{K}_v$ by construction, and

$$A(Av) = A^2 v = (\operatorname{trace} A)(Av) + (-\det A)v \in \mathcal{K}_v$$

due to the Cayley-Hamilton theorem. □

This simple observation will become useful in the next section.

4. Common invariant cones for families of 2 × 2 matrices

Let $\mathcal{A} = \{A_1, \ldots, A_n\}$ be a finite family of 2×2 real matrices. An $\mathcal{A}$-invariant proper cone by definition is A_j-invariant for all $j = 1, \ldots, n$, and in order for that to be possible each of the A_j's must be a Vandergraft matrix.

In particular, matrices of the form cI with $c < 0$ preclude the existence of $\mathcal{A}$-invariant proper cones. On the other hand, presence (or absence) of matrices cI with $c \geq 0$ in $\mathcal{A}$ is irrelevant. All such matrices (if any) can be deleted from $\mathcal{A}$ but may as well be left intact.

We first consider the case when all the matrices A_j share a dominant eigenvector u. If several of them are non-diagonalizable, we will say that they have the *same orientation* if the sets of vectors positively associated with u relative to these matrices coincide (of course, the sets of vectors negatively associated with u then coincide as well). This happens if and only if in a basis containing u the off diagonal elements of these matrices are all of the same sign.

Theorem 6. *Let $\mathcal{A} = \{A_1, \dots, A_n\}$ be a family of 2×2 Vandergraft matrices sharing the same dominant eigenvector u. Then there exists an $\mathcal{A}$-invariant proper cone $\mathcal{K}$ if and only if either*

(i) *all A_j are diagonalizable, and those of them (if any) which have $\det A_j < 0$, $\operatorname{trace} A_j = 0$ are scalar multiples of each other, or*

(ii) *all A_j have non-negative determinants, and those of them which are not diagonalizable (if any) have the same orientation.*

Proof. *"If" part.* (i) If all A_j are diagonalizable and have non-negative determinants, the result follows from Theorem 2: any proper cone $\mathcal{K}$ containing u and sufficiently narrow to avoid all the eigenvectors of A_j corresponding to their second eigenvalue will do the job.

Suppose now that some of A_j have negative determinants; relabel them by $A_1, \dots, A_k$. Consider $\mathcal{K} = \operatorname{Cone}\{u, v, A_1v, \dots, A_kv\}$, where v is a vector different from u but so close to it that $\mathcal{K}$ does not contain the non-dominant eigenvectors of $A_{k+1}, \dots, A_n$ and A_iA_j $(i, j = 1, \dots, k)$. The latter products are all Vandergraft matrices with positive determinants and, under conditions imposed, also diagonalizable (here, the hypothesis that all matrices A_j with negative determinants and zero traces are multiples of each other, is crucial). Hence, $\mathcal{K}$ is invariant under $A_{k+1}, \dots, A_n$ and A_iA_j $(i, j = 1, \dots, k)$ as in the previous part of the proof (see Theorem 2). In particular, $A_iA_1v, \dots, A_iA_kv \in \mathcal{K}$ for all $i = 1, \dots, k$. Since $A_iv \in \mathcal{K}$ $(i = 1, 2, \dots, k)$ by the construction of $\mathcal{K}$, in fact all the generators of $\mathcal{K}$ are mapped by $A_1, \dots, A_k$ into $\mathcal{K}$, so that $\mathcal{K}$ is invariant also under A_i for $i = 1, \dots, k$.

(ii) There is no need to consider the case when all A_j are diagonalizable, because it is covered by (i). Supposing that non-diagonalizable matrices are present in $\mathcal{A}$, relabel them by $A_1, \dots, A_k$. Choose a vector v positively associated with u relative to A_1; under the conditions imposed it will be positively associated with u also relative to $A_2, \dots, A_k$. By Theorem 3, $\mathcal{K} = \operatorname{Cone}\{u, v\}$ is A_j-invariant for $j = 1, \dots, k$. Moving v sufficiently close to u in order to avoid the non-dominant eigenvectors of $A_{k+1}, \dots, A_n$, we will make $\mathcal{K}$ invariant with respect to all $A_1, \dots, A_n$.

"Only if" part. In cases different from (i)–(ii) the family $\mathcal{A}$ contains either (iii) two linearly independent matrices with negative determinants and zero traces, or (iv) two non-diagonalizable matrices with different orientation, or (v) a non-diagonalizable matrix and a matrix with negative determinant.

Denote the matrices involved in each case by A_1 and A_2. Then in case (iii) A_1A_2 is a non-diagonalizable Vandergraft matrix, so that (iii) reduces to (v). In case (iv), due to the description given by Theorem 3 the intersection of any A_1-invariant proper cone with an A_2-invariant proper cone is a ray spanned by u, and therefore not proper. In case (v), the non-existence of common invariant proper cones follows from the comparison of Corollaries 1 and 2. $\square$

Corollary 3. *In the setting of Theorem* 6*, an $\mathcal{A}$-invariant proper cone exists if and only if any two matrices in the family $\mathcal{A}$ share an invariant proper cone.*

Proof. Indeed, from the consideration of cases (iii)–(v) in the proof of Theorem 6 it follows that there exists a pair of matrices in $\mathcal{A}$ with no common invariant proper cone, whenever conditions (i) or (ii) do not hold. $\square$

We now move to the situation when $\mathcal{A}$ contains matrices with different dominant eigenlines. As it happens, the crucial role is then played by an extended family $\mathcal{A}_1$ which contains $\mathcal{A}$ and all pairwise products (different from scalar multiples of the identity) of the matrices in $\mathcal{A}$ having negative determinants:

$$\mathcal{A}_1 = \mathcal{A} \cup \{A_iA_j \colon A_i, A_j \in \mathcal{A},\ \det A_i, \det A_j < 0 \text{ and } A_iA_j \neq cI\}.$$

Of course, $\mathcal{A}_1$ coincides with $\mathcal{A}$ if the latter consists only of matrices with non-negative determinants.

Theorem 7. *Let $\mathcal{A}$ be a finite family in $\mathbb{R}^{2\times 2}$. For an $\mathcal{A}$-invariant proper cone to exist it is necessary that*

(i) *all elements of $\mathcal{A}_1$ are Vandergraft matrices,*

(ii) *there are at most two dominant eigenlines corresponding to non-diagonalizable matrices in $\mathcal{A}_1$, and all of them (if there is more than one) corresponding to the same dominant eigenline also have the same orientation,*

and

(iii) *the dominant eigenlines of matrices in $\mathcal{A}_1$ are separated from the non-dominant ones.*

The *separation* condition (iii) means simply the existence of vectors v_1, v_2 such that the interior of $\text{Cone}\{v_1, v_2\}$ is free of the non-dominant eigenvectors of matrices in $\mathcal{A}_1$ while the interior of $\text{Cone}\{v_1, -v_2\}$ is free of the dominant eigenvectors of non-scalar matrices. The vectors v_j themselves are allowed to be both dominant and non-dominant, but only if as the latter they correspond to matrices in $\mathcal{A}_1$ with non-negative determinants.

Proof. An $\mathcal{A}$-invariant cone $\mathcal{K}$ also is $\mathcal{A}_1$-invariant. This immediately implies the necessity of condition (i).

According to Corollary 1, an eigenline of a non-diagonalizable 2×2 Vandergraft matrix must be passing through the boundary of any of its invariant proper cones. Thus, at most two such eigenlines are admissible.

If two non-diagonalizable matrices share the eigenline but have different orientation, the intersection of (any pair of) the respective invariant cones is an eigenray, due to Theorem 3, and therefore is not proper. These two observations settle the necessity of part (ii).

Finally, if $\mathcal{K}$ is an $\mathcal{A}$- (and therefore $\mathcal{A}_1$)-invariant proper cone, then all dominant eigenlines lie in $\mathcal{K} \cup (-\mathcal{K})$ while non-dominant eigenlines belong to the closure of the complement (to the complement itself, if the respective matrix has negative determinant). Thus, (iii) holds. $\square$

Suppose now that necessary conditions stated in Theorem 7 hold. Denote by $U = \{u_1, \ldots, u_N\}$ the set of all distinct dominant unit eigenvectors of matrices in $\mathcal{A}_1$ the directions of which are chosen in such a way that $\text{Cone}\, U$ is proper and its interior is free of non-dominant eigenlines (this is possible due to (iii)). If there are no such eigenlines (that is, all matrices in $\mathcal{A}$ are non-diagonalizable), impose instead the condition that u_j for $j = 2, \ldots, N$ are positively associated with u_1 relative to the matrix A_1 for which u_1 is an eigenvector (this is possible due to (ii)). This choice is unique up to changing the sign of all u_j simultaneously. Relabel also the elements of $\mathcal{A}$ in such a way that $\det A_i$ is negative for $i = 1, \ldots, k$ and non-negative otherwise (with the convention that $k = 0$ if $\det A_i \geq 0$ for all $i = 1, \ldots, n$).

For further consideration it is convenient to distinguish between the cases when there is none, one, or two dominant eigenlines corresponding to non-diagonalizable matrices in $\mathcal{A}_1$.

Theorem 8. *Let* $\mathcal{A} = \{A_1, \ldots, A_n\} \subset \mathbb{R}^{2 \times 2}$ *be such that all the elements of*

$\mathcal{A}_1$ *are diagonalizable matrices. Under the necessary conditions*[1] (i), (iii) *of Theorem 7 and using the notation introduced above, let*

$$\mathcal{K} = \mathrm{Cone}\{u_j, A_i u_j \colon i = 1, \ldots, k;\ j = 1, \ldots, N\}. \tag{6}$$

Then there exist $\mathcal{A}$-invariant proper cones if and only if the cone $\mathcal{K}$ is proper, its interior is free of the non-dominant eigenvectors of all matrices in $\mathcal{A}_1$, and the edges of $\mathcal{K}$ are not collinear with the eigenvectors of A_i ($i = 1, \ldots, k$).

Proof. "Only if" part. Any $\mathcal{A}$-invariant cone also is $\mathcal{A}_1$-invariant, and thus must contain either U or $-U$. Without loss of generality, let it contain U. Then, being invariant under all A_i, it must also contain $\mathcal{K}$. The rest follows from Theorems 2 and 4, applied to each of the matrices in $\mathcal{A}_1$.

"If" part. For $i = k+1, \ldots, n$, the cone $\mathcal{K}$ contains the dominant eigenvector of A_i (since it is one of the u_j's) and the interior of $\mathcal{K}$ does not contain its non-dominant eigenvectors. By Theorem 2, $\mathcal{K}$ is invariant under A_i.

Since $\det A_i A_m > 0$ for all $i, m = 1, \ldots, k$, the cone $\mathcal{K}$ for the same reasons is $A_i A_m$-invariant. Consequently, $A_i A_m u_j \in \mathcal{K}$ for all $i, m = 1, \ldots, k$; $j = 1, \ldots, N$. But $A_i u_j$ lies in $\mathcal{K}$ by construction. So, all the generators of $\mathcal{K}$ are mapped into $\mathcal{K}$ by $A_1, \ldots, A_n$. It remains to invoke Remark 1. $\square$

Theorem 8 shows that necessary conditions stated in Theorem 7 in general are not sufficient. For $k = 0$, however, $\mathcal{K}$ coincides with $\mathrm{Cone}\, U$, and the latter is $\mathcal{A}$-invariant already under the conditions of Theorem 7. The situation therefore simplifies as follows.

Corollary 4. *Let $\mathcal{A} = \{A_1, \ldots, A_n\}$ be a family of diagonalizable 2×2 matrices with non-negative determinants. Then in order for an $\mathcal{A}$-invariant proper cone to exist it is necessary and sufficient that*

(i) *all elements of $\mathcal{A}$ are Vandergraft matrices, and*

(ii) *the dominant eigenlines of matrices in $\mathcal{A}$ are separated from the non-dominant ones.*

We can now observe the following.

[1] Condition (ii) holds automatically.

Theorem 9. *Let $\mathcal{A} = \{A_1, \ldots, A_n\}$ be a family of diagonalizable 2×2 Vandergraft matrices with non-negative determinants. If any four of them (three — if there is at most one pair of simultaneously diagonalizable matrices in $\mathcal{A}$) have a common invariant proper cone, then there also exists an $\mathcal{A}$-invariant proper cone.*

Proof. Indeed, if an $\mathcal{A}$-invariant proper cone does not exist, then condition (iii) of Theorem 7 fails. But then it is possible to find four matrices in $\mathcal{A}$ (without loss of generality relabel them by $A_1, \ldots, A_4$) such that, when traveling around the origin in a counterclockwise direction, one encounters consequently the dominant eigenline of A_1, the non-dominant eigenline of A_2, the dominant eigenline of A_3, and finally the non-dominant eigenline of A_4. Condition (iii) fails for the set $\{A_1, A_2, A_3, A_4\}$, so that these four matrices already do not have a common invariant proper cone. Of course, it is not excluded that A_1 or A_3 coincides with A_2 or A_4, and then we have an even smaller subfamily of $\mathcal{A}$ with no common invariant proper cone. If A_1 and A_3 are not simultaneously diagonalizable, the non-dominant eigenline of at least one of them will be different from the dominant eigenline of the other. Consequently, in this case we can always choose A_2 or A_4 coinciding with A_1 or A_3. A similar reasoning works if a pair A_2, A_4 is not simultaneously diagonalizable. $\square$

We now move to the case of one dominant eigenline corresponding to non-diagonalizable matrices.

Theorem 10. *Let $\mathcal{A} = \{A_1, \ldots, A_n\} \subset \mathbb{R}^{2\times 2}$ satisfy conditions* (i)–(iii) *of Theorem* 7*, with exactly one dominant direction (say, corresponding to u_1) shared by all non-diagonalizable matrices in $\mathcal{A}_1$. Then there exists an $\mathcal{A}$-invariant proper cone if and only if the cone* (6) *satisfies conditions of Theorem* 8*, and in addition its interior consists of vectors positively associated with u_1.*

Proof. "Only if" part. An $\mathcal{A}$-invariant cone is invariant under all diagonalizable matrices in $\mathcal{A}_1$. Therefore, it must contain the cone (6). On the other hand, it is also invariant under all non-diagonalizable matrices in $\mathcal{A}_1$, so that by Theorem 3 u_1 must lie on its boundary, and the interior of the cone consists only of vectors positively associated with u_1. The same is therefore true for $\mathcal{K}$.

"If" part. As in Theorem 8, the cone (6) itself does the job. $\square$

The case of two dominant eigenlines corresponding to non-diagonalizable matrices in $\mathcal{A}_1$ can be treated along the same lines. However, a more straightforward (and less computationally consuming) approach also is available.

Suppose that conditions (i), (ii), and (iii) of Theorem 7 hold, and that $\mathcal{A}_1$ contains two non-diagonalizable matrices (say, B_1 and B_2) with non-collinear dominant eigenvectors. Relabel the latter as u_1 and u_2, choosing the direction of u_1 arbitrarily, and the direction of u_2 in such a way that it is positively associated with u_1 relative to B_1. According to Theorem 3, then either B_1 and B_2 have no common invariant proper cones (if u_1 is negatively associated with u_2 relative to B_2), or there are exactly two such cones: $\mathcal{K} = \mathrm{Cone}\{u_1, u_2\}$ and $-\mathcal{K}$.

Theorem 11. *For a finite family $\mathcal{A} = \{A_1, \dots, A_n\}$ of Vandergraft matrices with exactly two dominant eigenlines corresponding to non-diagonalizable matrices in $\mathcal{A}_1$, the only possible $\mathcal{A}$-invariant proper cones are $\pm\mathcal{K}$ introduced above. These cones are indeed $\mathcal{A}$-invariant if and only if:*

(a) *all non-diagonalizable matrices in $\mathcal{A}$ (if any) with a dominant eigenvector u_j have the same orientation as B_j ($j = 1, 2$),*

(b) *for all matrices $A_j \in \mathcal{A}$, their dominant eigenvectors lie in $\mathcal{K} \cup (-\mathcal{K})$ while the non-dominant ones lie outside the interior of $\mathcal{K} \cup (-\mathcal{K})$,*

(c) *for $A = A_j \in \mathcal{A}$ with the eigenvalues $\lambda_{1j} > 0$, $\lambda_{2j} < 0$ and the dominant eigenvector $u_1 + \xi u_2 \in \mathcal{K}$, the non-dominant eigenvector must be collinear with $u_1 + \eta u_2$, where*

$$\frac{\lambda_{1j}}{\lambda_{2j}} \le \frac{\eta}{\xi} \le \frac{\lambda_{2j}}{\lambda_{1j}}.$$

Proof. Indeed, conditions (a)–(c) are necessary and sufficient for $\mathcal{K}$ (or $-\mathcal{K}$) to be invariant under all matrices in $\mathcal{A}$, as follows by applying Theorem 2–4. And, as was observed earlier, no other proper cones can possibly be $\mathcal{A}$-invariant. $\square$

Remark 2. *It follows directly from the proof of Theorem 11 that if in its setting every three matrices in $\mathcal{A}_1$ (or any five matrices in $\mathcal{A}$) have a common invariant proper cone, then there also exists an $\mathcal{A}$-invariant proper cone.*

5. Simultaneously diagonalizable matrices

We now move to square matrices of arbitrary size $m \times m$ but suppose that all the elements of the family $\mathcal{A} = \{A_1, \ldots, A_n\}$ under consideration can be put in a diagonal form by the same similarity transformation S (note that S is allowed to be a complex matrix). This S then diagonalizes all matrices from $\mathcal{A}_2 = \operatorname{Cone} \mathcal{A}$, and moreover from the closed algebra $\mathcal{A}_3$ generated by $\mathcal{A}$. Denote by q ($\leq m$) the maximal number of distinct eigenvalues for matrices in $\mathcal{A}_2$. If B_0 is one of the matrices on which this number is attained,

$$B_0 = S \operatorname{diag}[b_1 I_{s_1}, \ldots, b_q I_{s_q}] S^{-1}, \qquad b_j \in \mathbb{C}, \quad b_i \neq b_j \text{ if } i \neq j, \tag{7}$$

then

$$A_j = S \operatorname{diag}[\lambda_{1j} I_{s_1}, \ldots, \lambda_{qj} I_{s_q}] S^{-1} \text{ for all } A_j \in \mathcal{A}; \text{ here } \lambda_{ij} \in \mathbb{C}. \tag{8}$$

Indeed, if at least one of the blocks in the middle factor of (8) were different from a scalar multiple of the identity, then the matrix $B_0 + \epsilon A_j$ would have more than q distinct eigenvalues for sufficiently small ϵ. (Note in passing, though this fact is not needed in what follows, that because of (8) q is also the maximal number of distinct eigenvalues of the matrices in a larger set $\mathcal{A}_3$.)

Assume there exists an $\mathcal{A}$-invariant proper cone. Then obviously all products $A_1^{m_1} \cdots A_n^{m_n}$ ($m_i \in \mathbb{Z}_+$, the set of nonnegative integers) are Vandergraft matrices. Due to the diagonalizability, this requirement amounts to $\max_i |\lambda_{i1}^{m_1} \cdots \lambda_{in}^{m_n}|$ being attained on some i for which $\lambda_{i1}^{m_1} \cdots \lambda_{in}^{m_n} \geq 0$. For every n-tuple $(m_1, \ldots, m_n) \in \mathbb{Z}_+^n$, introduce the set

$$\Omega(m_1, \ldots, m_n) = \{i_0 \in \{1, 2, \ldots, q\} \ : \ \max_{1 \leq i \leq q} |\lambda_{i1}^{m_1} \cdots \lambda_{in}^{m_n}| = \lambda_{i_0 1}^{m_1} \cdots \lambda_{i_0 n}^{m_n}\}.$$

Although $\Omega(m_1, \ldots, m_n)$ need not be a singleton, we note that there is a unique index $p = p(m_1, \ldots, m_n) \in \Omega(m_1, \ldots, m_n)$ for which

$$\max_{i_0 \in \Omega(m_1, \ldots, m_n)} |b_{i_0}| = b_p.$$

Indeed, this follows from the Vandergraft property of the matrix

$$A_1^{m_1} A_2^{m_2} \cdots A_n^{m_n} + B_0$$

and the condition $b_i \neq b_j$ if $i \neq j$. (Note that we take $X^0 = I$ for every square matrix X regardless if X is singular or not.) We let $P = \cup p(m_1, \ldots, m_n)$

where the union is taken over all n-tuples $(m_1, \dots, m_n) \in \mathbb{Z}_+^n$. Permuting the columns of S if necessary, we may suppose without loss of generality that this set is $P = \{1, \dots, k\}$, where $k \le q$.

Theorem 12. *In the notation* (8) *and for k as introduced above, $\mathcal{A}$-invariant proper cones exist if and only if*

$$\lambda_{ij} \ge 0 \text{ for all } i = 1, \dots k \text{ and } j = 1, \dots, n. \tag{9}$$

Proof. "Only if" part. For an arbitrarily fixed $i_0 \in \{1, \dots, k\}$, pick an n-tuple $m_1, \dots, m_n$ such that

$$\lambda_{i_0 1}^{m_1} \cdots \lambda_{i_0 n}^{m_n} = \max_{i=1,\dots q} |\lambda_{i1}^{m_1} \cdots \lambda_{in}^{m_n}| \,.$$

Then

$$\lambda_{i_0 1}^{m_1} \cdots \lambda_{i_0 n}^{m_n} + \epsilon > |\lambda_{i1}^{m_1} \cdots \lambda_{in}^{m_n} + \epsilon| \,, \quad i \notin \Omega(m_1, \dots, m_n)$$

for any $\epsilon > 0$, and therefore

$$\lambda_{i_0 1}^{m_1} \cdots \lambda_{i_0 n}^{m_n} + \epsilon + \delta b_{i_0} > |\lambda_{i1}^{m_1} \cdots \lambda_{in}^{m_n} + \epsilon + \delta b_i| \,, \quad i \ne i_0$$

for $\delta > 0$ small enough. Having fixed ϵ and δ (> 0), observe that then for any j such that $\lambda_{i_0 j} \ne 0$,

$$\left|(\lambda_{i_0 1}^{m_1} \cdots \lambda_{i_0 n}^{m_n} + \epsilon + \delta b_{i_0})^l \lambda_{i_0 j}\right| > \left|(\lambda_{i1}^{m_1} \cdots \lambda_{in}^{m_n} + \epsilon + \delta b_{i_0})^l \lambda_{ij}\right| , \quad i \ne i_0$$

if the positive integer l is large enough.

In other words, $(\lambda_{i_0 1}^{m_1} \cdots \lambda_{i_0 n}^{m_n} + \epsilon + \delta b_{i_0})^l \lambda_{i_0 j}$ is strictly bigger (by absolute value) than other eigenvalues of

$$B_l := (A_1^{m_1} \cdots A_n^{m_n} + \epsilon I + \delta B_0)^l A_j.$$

But an $\mathcal{A}$-invariant cone is also B_l-invariant whenever $\epsilon, \delta > 0$. So,

$$(\lambda_{i_0 1}^{m_1} \cdots \lambda_{i_0 n}^{m_n} + \epsilon + \delta b_{i_0})^l \lambda_{i_0 j} > 0.$$

Choosing two consecutive values of l, we conclude that in fact $\lambda_{i_0 j} > 0$.

"If" part. Denote by L_+ the (real) linear span of the first $s_1 + \dots + s_k$ columns of S. Note that since the eigenvalues of B_0 corresponding to these columns of S are real (see (7)), the first $s_1 + \dots + s_k$ columns of S are real as well (or more precisely can be made real if necessary, by (complex) scalings);

thus $L_+ \subset \mathbb{R}^m$. Let us represent $\mathbb{R}^m$ as the direct sum of the subspaces L_r and L_c spanned respectively by the real columns of S and by the real and imaginary parts of non-real (if any) columns of S. By definition of L_+, it lies in L_r. Moreover, L_r can be written as $L_r = L_+ \dot{+} L_-$, where L_- is also spanned by columns of S.

Choose bases $F_\pm$ in $L_\pm$ consisting of columns of S, and a basis F_c in L_c consisting of vectors $u_i, v_i \in \mathbb{R}^m$ such that

$$A_j u_i = (\operatorname{Re} \lambda_{ij}) u_i - (\operatorname{Im} \lambda_{ij}) v_i, \quad A_j v_i = (\operatorname{Im} \lambda_{ij}) u_i + (\operatorname{Re} \lambda_{ij}) v_i.$$

Then of course

$$\begin{aligned} (A_1^{m_1} A_2^{m_2} \cdots A_n^{m_n}) u_i &= (\operatorname{Re} \mu_i) u_i - (\operatorname{Im} \mu_i) v_i, \\ (A_1^{m_1} A_2^{m_2} \cdots A_n^{m_n}) v_i &= (\operatorname{Im} \mu_i) u_i + (\operatorname{Re} \mu_i) v_i, \end{aligned} \tag{10}$$

where $m_j \in \mathbb{Z}_+$ and $\mu_i = \lambda_{i1}^{m_1} \cdots \lambda_{in}^{m_n}$.

Denote by f the sum of all elements in F_+, and let $\mathcal{K}_0$ stand for the smallest $\mathcal{A}$-invariant convex cone containing F_+, $f + F_-$ and $f + F_c$. The span of $\mathcal{K}_0$ contains the basis $F = F_+ \cup F_- \cup F_c$ of the whole space $\mathbb{R}^m$, so that it coincides with $\mathbb{R}^m$. In other words, $\mathcal{K}_0$ is a *reproducing* convex cone, and therefore it is solid.

The closure $\mathcal{K}$ of $\mathcal{K}_0$ also is a convex solid cone invariant under $\mathcal{A}$. It remains only to show that $\mathcal{K}$ is pointed.

Let us relabel vectors in F by $f_1, \ldots, f_m$, with the first $p = s_1 + \cdots + s_k$ vectors belonging to F_+, and denote by $\alpha_j(v)$ the coordinates of the vector v in its expansion along F.

By (9), for $v = A_1^{m_1} \cdots A_n^{m_n} f_j$, $j = 1, \ldots, p$, we have

$$\alpha_j(v) \geq 0 \text{ and } \alpha_i(v) = 0 \text{ for all } i \neq j.$$

Consequently, for such v

$$\sum_{j=1}^{p} \alpha_j(v) \geq \sum_{p+1}^{m} |\alpha_j(v)|. \tag{11}$$

Inequality (11) obviously holds for $v \in f + F_-$ or $f + F_c$, since then the first p coordinates $\alpha_j(v)$ and exactly one of the other $m - p$ coordinates are equal to one, while the remaining ones are all zeros. The construction of the subspace L_+ (for which F_+ is a basis) guarantees that inequality (11) persists

for vectors v being images of $f + F_-$ under arbitrary products $A_1^{m_1} \cdots A_n^{m_n}$. Indeed, the left hand side of (11) is

$$\sum_{i=1}^{p} \lambda_{i1}^{m_1} \cdots \lambda_{im}^{m_n}, \tag{12}$$

while the right hand side is just one summand of the form

$$\left|\lambda_{j1}^{m_1} \cdots \lambda_{jm}^{m_n}\right|, \tag{13}$$

with j between $p+1$ and m. Since all summands in (12) are non-negative, and at least one of them is bigger than or equal to (13) — this is where the definition of L_+ is being used, — inequality (11) will hold for such v. Moreover, for images of $f + F_c$ under $A_1^{m_1} \cdots A_n^{m_n}$ we have, due to (10):

$$\sum_{j=1}^{p} \alpha_j(v) \geq \frac{1}{2} \sum_{p+1}^{m} |\alpha_j(v)|, \quad \alpha_j(v) \geq 0 \text{ for } j = 1, \ldots, p. \tag{14}$$

Since inequalities (11) and (14) persist under taking linear combinations with non-negative coefficients and passing to limits, we see that (11) holds in fact for all $v \in \mathcal{K}$. On the other hand, if (11) holds after switching from v to $-v$, then $\alpha_j(v) = 0$ for all $j = 1, \ldots, m$, so that $v = 0$. $\square$

6. Families of matrices with common dominant eigenvector

Theorem 6 gives a full treatment of families of 2×2 matrices sharing a dominant eigenvector. In higher dimensions, however, we have to impose additional restrictions.

Theorem 13. *Let $\mathcal{A}$ be a set of $m \times m$ Vandergraft matrices that share a common dominant eigenvector x and satisfy at least one of the following two conditions:*

(1) *The matrices in $\mathcal{A}$ are simultaneously similar, with a real similarity matrix, to normal matrices;*

(2) *$\mathcal{A}$ is finite, the matrices in $\mathcal{A}$ commute and for every $A \in \mathcal{A}$, $\rho(A)$ is a semisimple eigenvalue, i.e., a simple root of the minimal polynomial, of A.*

Then the matrices in $\mathcal{A}$ have a common invariant proper cone $\mathcal{K}$ with the additional property that x belongs to the interior of $\mathcal{K}$.

For the proof of Theorem 13 we need two lemmas.

Lemma 14. *Let $A_1, \ldots, A_q$ be commuting $m \times m$ real matrices. Assume that there exists λ_0 real with the following properties:*

(1) *there exists a nonzero x such that $A_j x = \lambda_0 x$ for $j = 1, 2, \ldots, q$.*

(2) *λ_0 is a semisimple eigenvalue of A_j, for $j = 1, 2, \ldots, q$.*

Then there exists an invertible real matrix S such that $S^{-1} A_j S$ have the form

$$S^{-1} A_j S = \begin{bmatrix} \lambda_0 & 0 \\ 0 & B_j \end{bmatrix}, \quad j = 1, 2, \ldots, q,$$

where $B_1, \ldots, B_q$ are $(m-1) \times (m-1)$ matrices.

Proof. Induction on q. For $q = 1$, the result is clear. Assume Lemma 14 has been proved for $q - 1$ matrices. Applying a simultaneous similarity to $A_1, \ldots, A_q$, we may assume that

$$A_1 = \begin{bmatrix} \lambda_0 I_p & 0 \\ 0 & \widetilde{A}_1 \end{bmatrix},$$

where λ_0 is not an eigenvalue of $\widetilde{A}_1$. Since $A_1, \ldots, A_q$ commute we have

$$A_j = \begin{bmatrix} B_j & 0 \\ 0 & C_j \end{bmatrix}, \quad j = 2, 3, \ldots, q.$$

Here the matrices $B_2, \ldots, B_q$ are $p \times p$. Clearly, the vector x (which exists by (1)) has the form $x = \begin{bmatrix} y \\ 0 \end{bmatrix}$, where $y \neq 0$ has p components. Then $B_j y = \lambda_0 y$. One verifies that λ_0 is a semisimple eigenvalue of each B_j. By the induction hypothesis, there exists an invertible real T such that

$$T^{-1} B_j T = \begin{bmatrix} \lambda_0 & 0 \\ 0 & \widetilde{B}_j \end{bmatrix}, \quad j = 2, 3, \ldots, q.$$

Now take $S = \begin{bmatrix} T & 0 \\ 0 & I \end{bmatrix}$ to satisfy the lemma. □

Lemma 15. *Let $A_1, \dots, A_q$ be commuting $m \times m$ complex matrices with the following properties:*

(1) $\rho(A_j) \le 1$ *for* $j = 1, 2, \dots, q$*;*

(2) *every eigenvalue (if exists) on the unit circle of every A_j is semisimple.*

Then there exists a positive definite matrix V such that

$$V - A_j^* V A_j \ge 0, \quad \text{for } j = 1, 2, \dots, q. \tag{15}$$

($A \ge B$ means that $A - B$ is positive semidefinite).

Moreover, if all A_j's are real, then V can be also chosen real.

Proof. It is enough to prove the complex case only. Indeed, suppose all A_j's are real and we have proved that there exists a (generally, complex) positive definite V such that (15) holds. Then by taking complex conjugates in (15) we obtain

$$\overline{V} - A_j^T \overline{V} A_j \ge 0, \quad j = 1, 2, \dots, j = 1, 2, \dots, q. \tag{16}$$

Adding (15) and (16) we see that $U - A_j^T U A_j \ge 0$, where $U := V + \overline{V}$ is positive definite and real.

We now prove the complex case. If $\rho(A_j) < 1$ for all j, let

$$V = \sum (A_1^*)^{z_1} \cdots (A_q^*)^{z_q} A_q^{z_q} \cdots A_1^{z_1}, \tag{17}$$

where the sum is taken over all q-tuples $(z_1, \dots, z_q)$, $z_j \in \mathbb{Z}_+$. It is easy to see (using $\rho(A_j) < 1$) that the series in (17) converges absolutely. Clearly $V \ge I$ and

$$\begin{aligned} V \ &- \ A_j^* V A_j \\ &= \ \sum (A_1^*)^{z_1} \cdots (A_{j-1}^*)^{z_{j-1}} (A_{j+1}^*)^{z_{j+1}} \cdots (A_q^*)^{z_q} A_q^{z_q} \cdots A_{j+1}^{z_{j+1}} A_{j-1}^{z_{j-1}} \cdots A_1^{z_1} \\ &\ge \ 0, \end{aligned}$$

where the sum is taken over all $(q-1)$-tuples $(z_1, \dots, z_{j-1}, z_{j+1}, \dots, z_q) \in \mathbb{Z}_+^{q-1}$.

So suppose that $\rho(A_j) = 1$ for some j, say $\rho(A_1) = 1$. Note that the hypotheses and the conclusions of Lemma 15 are invariant under simultaneous similarity of $A_1, \dots, A_q$:

$$A_j \mapsto S^{-1} A_j S, \quad j = 1, 2, \dots, q,$$

where S is any invertible $m \times m$ matrix. Then, considering each root subspace of A_1 separately, and taking advantage of the commutativity property $A_j A_k = A_k A_j$ for $j, k = 1, 2, \ldots, q$, we reduce the proof to the case $A_1 = \lambda I$, $|\lambda| = 1$. Then obviously $V - A_1^* V A_1 = 0$, and it suffices to prove (15) for $A_2, \ldots, A_q$. This follows by induction on q, the case $q = 1$ being easy. $\square$

We now proceed with the proof of Theorem 13.

Proof. Assume first that (1) holds. We may assume that $\mathcal{A}$ consists of normal matrices and that $\|x\| = 1$ (the norm is Euclidean). Let $\mathbb{M}$ be the orthogonal complement to Span $\{x\}$. We claim that:

$$\mathcal{K} := \{c_1 x + y \, : \, c_1 \in \mathbb{R}, \quad y \in \mathbb{M}, \quad c_1 \geq \|y\|\}$$

is a common invariant proper cone for all $A \in \mathcal{A}$.

Clearly, $\mathcal{K}$ is a proper cone; therefore we only have to show that it is invariant with respect to the matrices. Let $A \in \mathcal{A}$, and let $x, u_2, \ldots, u_m$ be an orthonormal set with the following properties:

$$Au_{2k} = \alpha_k u_{2k} + \beta_k u_{2k+1}, \quad Au_{2k+1} = -\beta_k u_{2k} + \alpha_k u_{2k+1}, \quad \text{for } \ k = 1, 2, \ldots, \ell,$$

$$Au_s = \lambda_s u_s \quad \text{for } \ s = 2\ell + 2, 2\ell + 3, \ldots, m,$$

where $\alpha_k, \beta_k, \lambda_s$ are real numbers such that

$$\beta_k > 0, \quad |\lambda_s| \leq \rho(A), \quad \sqrt{\alpha_k^2 + \beta_k^2} \leq \rho(A);$$

here ℓ is a certain nonnegative integer. (The existence of such $u_2, \ldots, u_m$ follows from the canonical form of real normal matrices, see, e.g., [13].) Obviously $u_2, \ldots, u_m$ form an orthonormal basis in $\mathbb{M}$. Take

$$y = c_1 x + c_2 u_2 + \ldots + c_m u_m \in \mathcal{K},$$

thus $c_1 \geq \sqrt{c_2^2 + \cdots + c_m^2}$. Then we have

$$\begin{aligned} Ay = \rho(A) c_1 x + c_2(\alpha_1 u_2 + \beta_1 u_3) + c_3(-\beta_1 u_2 + \alpha_1 u_3) + \cdots \\ + c_{2\ell}(\alpha_\ell u_{2\ell} + \beta_\ell u_{2\ell+1}) + c_{2\ell+1}(-\beta_\ell u_{2\ell} + \alpha_\ell u_{2\ell+1}) + \\ \lambda_{2\ell+2} c_{2\ell+2} u_{2\ell+2} + \ldots + \lambda_m c_m u_m := \rho(A) c_1 x + w. \end{aligned} \tag{18}$$

Notice that for $k = 1, 2, \ldots, \ell$ we have

$$\begin{aligned} c_{2k}(\alpha_k u_{2k} + \beta_k u_{2k+1}) \quad &+ \quad c_{2k+1}(-\beta_k u_{2k} + \alpha_k u_{2k+1}) \\ &= \quad (c_{2k}\alpha_k - c_{2k+1}\beta_k)u_{2k} \quad + \quad (c_{2k}\beta_k + c_{2k+1}\alpha_k)u_{2k+1} \end{aligned}$$

and

$$\frac{(c_{2k}\alpha_k - c_{2k+1}\beta_k)^2 + (c_{2k}\beta_k + c_{2k+1}\alpha_k)^2}{\rho(A)^2} = \frac{(\alpha^2+\beta^2)(c_{2k}^2 + c_{2k+2}^2)}{\rho(A)^2} \leq c_{2k}^2 + c_{2k+1}^2.$$

Thus,

$$\|w/\rho(A)\|^2 \leq c_2^2 + \cdots + c_{2\ell+1}^2 + \frac{\lambda_{2\ell+2}^2}{\rho(A)^2} c_{2\ell+2}^2 + \cdots + \frac{\lambda_m^2}{\rho(A)^2} c_m^2 \leq c_2^2 + \cdots + c_m^2, \quad (19)$$

and it follows from (18) and (19) that $Ay \in \mathcal{K}$.

Assume now that (2) of Theorem 13 holds. Let $\mathcal{A} = \{A_1, \ldots, A_q\}$. We may assume that the spectral radius of each A_j is positive (if some A_j is nilpotent, the hypotheses of Theorem 13 (assuming (2)) imply that it is actually equal to the zero matrix, and can be ignored). Scaling the A_j's we may further assume that $\rho(A_j) = 1$, $j = 1, 2, \ldots, q$. By Lemma 14 we may assume that

$$A_j = \begin{bmatrix} 1 & 0 \\ 0 & B_j \end{bmatrix},$$

where $B_1, \ldots, B_q$ are $(m-1) \times (m-1)$ matrices. By Theorem 1, the hypotheses (1) and (2) of Lemma 15 are satisfied for $B_1, \ldots, B_q$. Thus, there exists a real positive definite matrix V such that

$$V - B_j^T V B_j \geq 0, \quad j = 1, 2, \ldots, q. \quad (20)$$

Then

$$\mathcal{K} := \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x \geq 0, \ \ y \in \mathbb{R}^{m-1} \text{ is such that } y^T V y \leq x^2 \right\}$$

is a common invariant cone for $A_1, \ldots, A_q$. Indeed, if $\begin{bmatrix} x \\ y \end{bmatrix} \in \mathcal{K}$, then

$$A_j \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ B_j y \end{bmatrix},$$

and

$$(B_j y)^T V B_j y \leq \text{ by (20) } \leq y^T V y \leq x^2,$$

and so

$$A_j \begin{bmatrix} x \\ y \end{bmatrix} \in \mathcal{K}.$$

Clearly, $\mathcal{K}$ is topologically closed, is closed under multiplication by nonnegative real numbers, is solid and pointed, because of the positive definiteness of V. It remains to prove that $\mathcal{K}$ is convex. Thus, let $x_1, x_2 \geq 0$ and $y_1, y_2 \in \mathbb{R}^{m-1}$ be such that

$$y_k^T V y_k \leq x_k^2, \quad \text{for } k = 1, 2. \tag{21}$$

Then for a number α between 0 and 1, we have:

$$(\alpha y_1 + (1-\alpha) y_2)^T V (\alpha y_1 + (1-\alpha) y_2) \leq \alpha^2 x_1^2 + (1-\alpha)^2 x_2^2 + 2\alpha(1-\alpha)(y_1^T V y_2) \leq$$

$$\leq \alpha^2 x_1^2 + (1-\alpha)^2 x_2^2 + 2\alpha(1-\alpha) x_1 x_2 = (\alpha x_1 + (1-\alpha) x_2)^2$$

(Cauchy-Schwartz inequality and (21) are used in the last step of the derivation), and the convexity of $\mathcal{K}$ is proved. $\square$

7. Examples

In this section we collect examples that illuminate concepts and results presented. We use the notation

$$\mathbf{e}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad \mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Example 1. *Two 2×2 matrices A and B with negative determinants such that all words in A and B are Vandergraft matrices though there is no (A, B)-invariant proper cone.*

Take
$$A = \begin{bmatrix} 1 & p \\ 0 & -1 \end{bmatrix}, \; B = \begin{bmatrix} 1 & q \\ 0 & -1 \end{bmatrix}, \quad p \neq q.$$

All words in A and B are Vandergraft matrices, with $u_1 = \mathbf{e}_1$ as a dominant eigenvector. So, Theorem 6 applies, and according to case (iii) in "Only if" part of its proof (A, B)-invariant proper cones do not exist.

Example 1 shows that Theorem 7.6 in [11] is apparently misstated.

Example 2. *A triple of matrices* $T := \{A, B, C\}, \quad A, B, C \in \mathbb{R}_V^{2\times 2}$ *with the following properties:*

(a) $\det M > 0$ *for all* $M \in T$*;*

(b) A, B, C *are normal matrices (in particular, diagonalizable);*

(c) *there is no* T*-invariant proper cone;*

(d) *each pair of matrices in* T *has a common invariant proper cone;*

(e) *no two matrices in* T *have a common eigenvector.*

The example shows that sharing a common dominant eigenvector is essential in Corollary 3 and Theorem 13, and also that the part of Theorem 9 pertinent to the case when there are no simultaneously diagonalizable pairs of matrices in $\mathcal{A}$ is sharp.

Instead of describing the matrices directly, we will list two linearly independent eigenvectors and associated eigenvalues for each matrix. For the eigenvalues simply pick $\lambda_1(M) > \lambda_2(M) > 0$ for each matrix $M \in T$. As for the eigenvectors of a matrix M, denoting the dominant and non-dominant ones by $u_1(M)$ and $u_2(M)$ respectively, let

$$u_1(A) = \mathbf{e}_1,\ u_1(B) = \begin{bmatrix} 1 \\ 2 \end{bmatrix},\ u_1(C) = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$$

and

$$u_2(A) = \mathbf{e}_2,\ u_2(B) = \begin{bmatrix} -2 \\ 1 \end{bmatrix},\ u_2(C) = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Each of the pairs (A, B), (A, C) and (B, C) then satisfies conditions of Corollary 4, and therefore has a common invariant proper cone (more specifically, $\mathrm{Cone}\{u_1(A), u_1(B)\}$ is (A, B)-invariant, $\mathrm{Cone}\{-u_1(B), u_1(C)\}$ is (B, C)-invariant, and $\mathrm{Cone}\{u_1(A), u_1(C)\}$ is (A, C)-invariant). On the other hand, the separation condition (ii) of Corollary 4 does not hold for the triple (A, B, C), so that there is no (A, B, C)-invariant proper cone.

Example 3. *A quadruple of matrices* $A, B, C, D \in \mathbb{R}^{2\times 2}$ *with distinct positive eigenvalues such that each triple of them has a common invariant proper cone while there is no* (A, B, C, D)*-invariant proper cone.*

In accordance with Theorem 9, this quadruple consists of two pairs of commuting matrices.

As in Example 2, the eigenvalues of the matrices can be chosen arbitrarily, as long as they are positive and distinct. Following the eigenvector notation from the same Example, let

$$u_1(A) = u_2(B) = \mathbf{e}_1,\ u_2(A) = u_1(B) = \mathbf{e}_2,$$

$$u_1(C) = u_2(D) = \mathbf{e}_1 + \mathbf{e}_2,\ u_2(C) = u_1(D) = \mathbf{e}_1 - \mathbf{e}_2.$$

The vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_1 + \mathbf{e}_2, \mathbf{e}_1 - \mathbf{e}_2$ are simultaneously dominant and non-dominant for the quadruple (A, B, C, D), and cannot be separated in the sense of condition (iii) of Theorem 7. Consequently, there is no (A, B, C, D)-invariant proper cone. On the other hand, from Corollary 4 it follows (and can also be checked directly, based on Theorem 2) that $\mathrm{Cone}\{\mathbf{e}_1, \mathbf{e}_2\}$ is (A, B, C)-invariant, $\mathrm{Cone}\{\mathbf{e}_1, -\mathbf{e}_2\}$ is (A, B, D)-invariant, $\mathrm{Cone}\{\mathbf{e}_1 + \mathbf{e}_2, \mathbf{e}_1 - \mathbf{e}_2\}$ is (A, C, D)-invariant, and $\mathrm{Cone}\{\mathbf{e}_1 + \mathbf{e}_2, \mathbf{e}_2 - \mathbf{e}_1\}$ is (B, C, D)-invariant.

Example 4. *The set $S = \{A, B\}$ which satisfies all the hypotheses of Theorem 13 (with (2) holding) except that $\rho(A)$, $\rho(B)$ are not semisimple eigenvalues of A, B, respectively, and there is no (A, B)-invariant proper cone.*

Take

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad , \quad B = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}.$$

Clearly, both matrices are Vandergraft, non-diagonalizable, sharing the dominant eigenline but having different orientation. By Theorem 7, there is no common invariant proper cone.

Example 5. *Two diagonal matrices A_1 and A_2 without a common invariant proper cone such that all words in A_1 and A_2 are Vandergraft matrices:*

$$A_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

It is easy to check that all words in A_1 and A_2 are Vandergraft matrices. However, condition (9) of Theorem 12 fails, so that there is no (A_1, A_2)-invariant proper cone.

Example 6. *Countable set of 2×2 Vandergraft matrices such that every finite number of them has a common invariant proper cone, but the whole set does not:*

Using Theorem 2, it is easy to see that any set of the form

$$\left\{ A_m = \begin{bmatrix} 1 & q_m \\ 0 & r \end{bmatrix}, \qquad m = 1, 2, \ldots, \right\},$$

where the sequence $\{|q_m|\}_{m=1}^{\infty}$ tends to infinity and $0 \leq r < 1$ is fixed, fits the bill.

Remark 3. *From standard compactness considerations it follows that if $\mathcal{A}$ is an infinite family in $\mathbb{R}_V^{n \times n}$ any finite subfamily of which has a common invariant proper cone, then there exists a non-trivial (that is, different from $\{0\}$) $\mathcal{A}$-invariant closed convex pointed cone. However, it may not be solid, and therefore is not necessarily proper.*

This is exactly what is happening in Example 6.

References

[1] A. Berman, R. J. Plemmons, Nonnegative matrices in the mathematical sciences, SIAM, Philadelphia, PA, 1994, revised reprint of the 1979 original.

[2] R. B. Bapat, T. E. S. Raghavan, Nonnegative matrices and applications, Cambridge University Press, Cambridge, 1997.

[3] B.-S. Tam, A cone-theoretic approach to the spectral theory of positive linear operators: the finite-dimensional case, Taiwanese J. Math. 5 (2) (2001) 207–277.

[4] G. Birkhoff, Linear transformations with invariant cones, Amer. Math. Monthly 74 (1967) 274–276.

[5] J. S. Vandergraft, Spectral properties of matrices which have invariant cones, SIAM J. Appl. Math. 16 (1968) 1208–1222.

[6] B.-S. Tam, H. Schneider, On the core of a cone-preserving map, Trans. Amer. Math. Soc. 343 (2) (1994) 479–524.

[7] B.-S. Tam, The Perron generalized eigenspace and the spectral cone of a cone-preserving map, Linear Algebra Appl. 393 (2004) 375–429.

[8] G. P. Barker, On matrices having an invariant cone, Czechoslovak Math. J. 22(97) (1972) 49–68.

[9] M. E. Valcher, L. Farina, An algebraic approach to the construction of polyhedral invariant cones, SIAM J. Matrix Anal. Appl. 22 (2) (2000) 453–471.

[10] A. Tiwari, J. Fung, Polyhedral cone invariance applied to rendezvous of multiple agents, in: 43rd IEEE Conference on Decision and Control, Vol. 1, Dec. 2004, pp. 165–170.

[11] R. Edwards, J. J. McDonald, M. J. Tsatsomeros, On matrices with common invariant cones with applications in neural and gene networks, Linear Algebra Appl. 398 (2005) 37–67.

[12] V. D. Blondel, Y. Nesterov, Computationally efficient approximations of the joint spectral radius, SIAM J. Matrix Anal. Appl. 27 (1) (2005) 256–272.

[13] R. A. Horn, C. R. Johnson, Matrix Analysis, Cambridge University Press, New York, 1985.